



Dianda, Daniela
Hernández, Lucia
Quaglino, Marta
Pagura, José Alberto

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística

INDICADORES MULTIVARIADOS DE CAPACIDAD DE PROCESOS. SU EFICIENCIA BAJO DISTRIBUCIONES NORMALES.

Introducción.

El avance de las técnicas de análisis multivariado de datos producido durante las últimas décadas, ha hecho posible que se produjera una transición desde el control univariado de los procesos hacia la consideración de diversas variables de calidad de manera simultánea, proveyendo las herramientas necesarias para tratar al problema precisamente como un problema multivariado. La ventaja de este tratamiento frente al enfoque usual de analizar por separado cada una de las características de calidad de interés, es que los nuevos procedimientos propuestos tienen en cuenta la estructura de correlación entre las variables de calidad. Estas propuestas, según el enfoque que han considerado para construir los índices, pueden clasificarse en cuatro categorías: 1) índices basados en cociente de volúmenes, entre la región de variación natural del proceso y la región de especificación, 2) índices definidos haciendo uso del análisis de componentes principales, 3) índices basados en las probabilidades de producir productos fuera de especificación, y 4) otros enfoques diversos, tales como los índices basados en la función de pérdida, en extensiones de los índices univariados y las propuestas de vectores de capacidad multidimensionales.

En este trabajo, se enfocan comparativamente algunas de las alternativas propuestas en las dos primeras categorías. Entre los índices basados en cociente de volúmenes de regiones, se consideran la propuesta inicial debida a W. Taam et al. (1993), el índice MC_{pm} , y una propuesta reciente debida a H. Shahriari et al. (2009), el índice NMC_{pm} . Entre los índices definidos mediante componentes principales se considera el propuesto inicialmente por F. K. Wang et al. (1998) y la propuesta de R. L. Shinde et al. (2008) que sugiere una mejora al índice de Wang. Se lleva a cabo un estudio por simulación cuyo objetivo es evaluar la capacidad de estos indicadores para identificar distintas situaciones de alejamiento del producto respecto de las especificaciones. Las comparaciones del comportamiento de cada



índice se llevan a cabo frente a diferentes escenarios que combinan número de variables, intensidad de las asociaciones entre ellas, tamaños de muestra y capacidad del proceso.

Metodología.

Sea X un vector que representa las mediciones de p características de calidad, provenientes de una distribución normal con vector de medias μ y matriz de variancias y covariancias Σ . Sean LIE , LSE y T los vectores de dimensión $p \times 1$ que contienen, respectivamente, los límites de especificación inferior y superior y los valores objetivos de cada una de las p variables de calidad.

El índice propuesto por Taam et al. (1993) se define como el cociente entre el volumen de la región de tolerancia relativo al volumen de la región de variación natural del proceso.

Las especificaciones impuestas a las características de calidad generan en el espacio p -dimensional de las variables una región hiper-rectangular, mientras que si el proceso subyacente sigue una distribución normal multivariada, la región de variación natural de los datos tendrá la forma de un hiper-elipsoide.

En orden de realizar una comparación "razonable" entre los volúmenes de ambas regiones, se propone modificar alguna de ellas para que ambas sean similares en forma. La propuesta de este autor sugiere modificar la región de tolerancia hiper-rectangular, considerando el mayor hiper-elipsoide centrado en el objetivo, de ejes paralelos a los ejes coordenados en el espacio de las variables y completamente contenido en la región de tolerancia original (Figura 1). El índice se calcula como:

$$MC_{pm} = \frac{Vol R_1}{Vol R_2} = \frac{Vol \text{ región de tolerancia modificada}}{Vol \{X - \mu' \Sigma_T^{-1} X - \mu \leq \chi_{1-\alpha, p}^2\}} \quad (1)$$

Donde R_1 es la región de tolerancia modificada, y R_2 es la región de variación natural del proceso que contiene un determinado porcentaje de los datos (comúnmente, el 99.73%). En esta última región, la matriz de la forma cuadrática es la matriz del error cuadrático medio del proceso, $\Sigma_T = E \{X - T \{X - T\}'$, de modo que el índice es una medida real de la capacidad del proceso, y no sólo potencial. En efecto, es posible demostrar que el índice puede escribirse como el producto de dos componentes, una que considera la variabilidad del proceso respecto a las especificaciones, y otra que considera las desviaciones del centrado del proceso respecto del objetivo.



Posteriormente a la propuesta de Taam, varios autores sugirieron modificaciones sobre el indicador MC_{pm} , argumentando principalmente que tal indicador ignora por completo la estructura de correlación de los datos cuando modifica la región de tolerancia. Entre estas sugerencias, se considera la realizada por los autores Shahriari et al. (2009), quienes propusieron considerar como región de tolerancia modificada al mayor hiper-elipsoide centrado en el objetivo, completamente contenido en la región de tolerancia, pero cuyos ejes sean paralelos a los ejes del hiper-elipsoide que define a la región de variación natural del proceso (Figura 2). El índice propuesto por estos autores, NMC_{pm} , se calcula como en (1), pero teniendo en cuenta para el numerador la nueva región de tolerancia modificada.

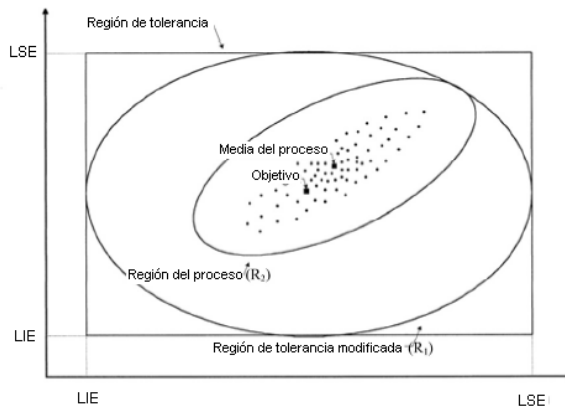


Figura 1: Región de tolerancia modificada según la propuesta de Taam et al. para $m=2$.

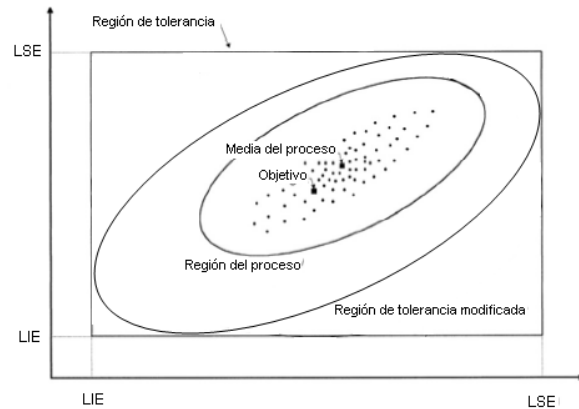


Figura 2: Región de tolerancia modificada según la propuesta de Shahriari et al. para $m=2$.

Dentro de la línea de índices construidos haciendo uso del análisis de componentes principales, la propuesta de los autores Wang y Chen en el año 1998, básicamente consiste en aplicar componentes principales al conjunto de p variables de calidad de interés, de modo de obtener un nuevo conjunto de variables (las componentes principales, Y_i , $i = 1, \dots, p$) independientes y con distribución normal: $Y_i = \mathbf{u}_i' \mathbf{X}$, donde \mathbf{u}_i es el i -ésimo autovector de la matriz Σ . Esta transformación, aplicada también a los vectores que contienen los límites de especificación y valores objetivos, ($LIE_{PC_i} = \mathbf{u}_i' LIE$, $LSE_{PC_i} = \mathbf{u}_i' LSE$, $T_{PC_i} = \mathbf{u}_i' T$) permiten "trasladar" el problema a un nuevo sistema de coordenadas definido por las componentes principales. En este nuevo contexto, el índice de capacidad propuesto se define simplemente como el producto de los índices univariados para cada una de las Y_i :



$$MC_{pm} = \prod_{i=1}^p C_{pm,Y_i}^{1/p}$$

donde C_{pm,Y_i} es el índice univariado C_{pm} calculado para la i -ésima componente principal. Adicionalmente, C_{pm,Y_i} puede reemplazarse por las opciones C_{p,Y_i} , C_{pk,Y_i} o C_{pmk,Y_i} , para producir versiones multivariadas de los índices univariados más tradicionales.

Los autores Shinde y Khadse, en el año 2009, realizaron una crítica a este método argumentando que el mismo presenta un gran problema en la identificación de la región de especificación en el espacio de las componentes principales. Con el método originalmente propuesto, la obtención de la nueva región de especificación en el espacio de las componentes principales asume que los límites de especificación para las Y_i son independientes entre sí, hecho que no es correcto. En efecto, si la región de especificación inicial es el hiper-rectángulo $S = \mathbf{x} : LIE \leq \mathbf{x} \leq LSE$, entonces la región de especificación para las nuevas variables \mathbf{Y}_i corresponde a una región en el espacio de las componentes principales definida por $S_2 = \mathbf{y} : LIE \leq \mathbf{U}'\mathbf{y} \leq LSE$, donde $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p)$, (Figuras 3 y 4).

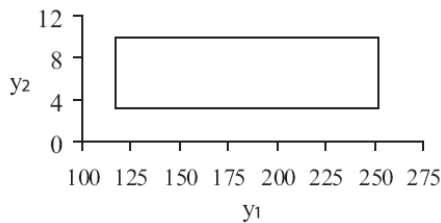


Figura 3: Región de tolerancia para las CPs según propuesta de Wang para $m=2$.

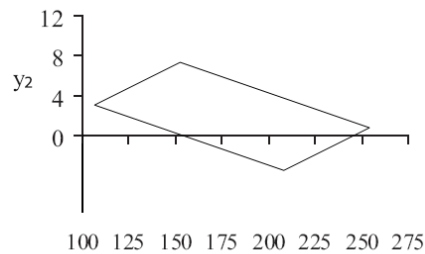


Figura 4: Región de tolerancia para las CPs según propuesta de Shinde et al. para $m=2$.

Dada la complejidad que pueden tener asociadas las regiones de especificación modificadas así definidas, el índice de capacidad propuesto por estos autores se define en términos de la proporción de productos "conformes", es decir, que caen dentro de la región de especificación modificada, y se estima mediante un enfoque empírico, haciendo uso de procedimientos de simulación.



Resultados.

Intensos estudios de simulación¹ permitieron evaluar el comportamiento de los índices considerados frente a una amplia variedad de situaciones. Los procesos de simulación fueron diseñados de modo de poder realizar las comparaciones en dos sentidos: comparar el comportamiento de cada índice frente a escenarios diferentes, con el objetivo de evaluar la bondad de la medida para señalar correctamente el estado real del proceso al variar los factores considerados; y comparar como responden los diferentes índices frente a un escenario particular.

Se presentan a continuación los resultados obtenidos para los cuatro índices estudiados en una de las situaciones particulares ensayadas. El caso seleccionado corresponde al estudio de muestras de tamaño 60 de un proceso capaz tanto en centrado como dispersión, proveniente de una distribución normal bivariada pura, donde se ha variado el coeficiente de correlación entre las variables de 0 a 0.95 en pasos de 0.05. Los resultados se muestran en la tabla 1.

Tabla 1: Valores medios y desvíos estándar de los índices de capacidad, calculados sobre 20.000 simulaciones.

ρ	Taam		Shahriari		Wang		Shinde ^(*)	
	Media	SD	Media	SD	Media	SD	Media	SD
0	1,0535	0,1395	0,9601	0,0711	0,8933	0,2820	0,9992	0,0012
0.10	1,0571	0,1412	0,9606	0,0718	0,8452	0,2931	0,9993	0,0011
0.20	1,0760	0,1419	0,9605	0,0719	0,7433	0,2955	0,9993	0,0011
0.30	1,1053	0,1460	0,9622	0,0724	0,6369	0,2748	0,9993	0,0012
0.40	1,1514	0,1529	0,9644	0,0741	0,5565	0,2432	0,9993	0,0011
0.50	1,2178	0,1608	0,9671	0,0757	0,4970	0,2187	0,9993	0,0011
0.60	1,3169	0,1735	0,9709	0,0783	0,4528	0,1982	0,9993	0,0011
0.70	1,4733	0,1943	0,9742	0,0797	0,4168	0,1837	0,9992	0,0013
0.80	1,7573	0,2342	0,9808	0,0837	0,3938	0,1730	0,9991	0,0012
0.90	2,4160	0,3181	0,9902	0,0882	0,3712	0,1625	0,9991	0,0013

(*) Estos valores representan proporciones de productos conformes.

La figura 5 muestra la representación gráfica de estos resultados para los tres primeros indicadores. Se puede observar que efectivamente la propuesta inicial basada en cociente de volúmenes de regiones, presenta algún tipo de inconveniente para reflejar de manera correcta el estado real del proceso. El índice propuesto por Taam, se ve claramente

¹ Los estudios por simulación se realizaron utilizando el software estadístico R, utilizando como base algunas de las funciones del paquete MPCl (Santos Fernández (2012)).



afectado por variaciones en la correlación de las variables, generando una sobreestimación de la capacidad a medida que la correlación entre las características aumenta. La modificación propuesta por Shahriari sobre el índice de Taam efectivamente parece corregir el inconveniente, resultando en valores de índices mucho más estables ante variaciones en la correlación. Las figuras 6 (a) y (b) muestran el efecto que produce el hecho de no tener en cuenta la asociación entre las variables al modificar la región de especificación.

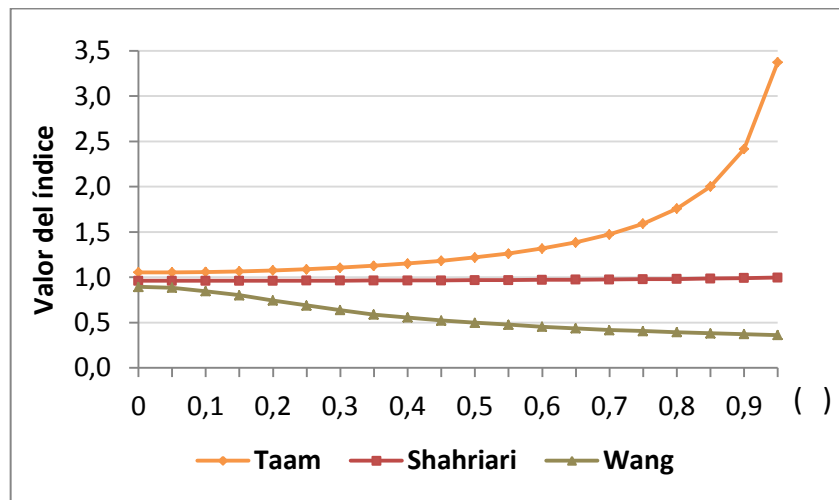
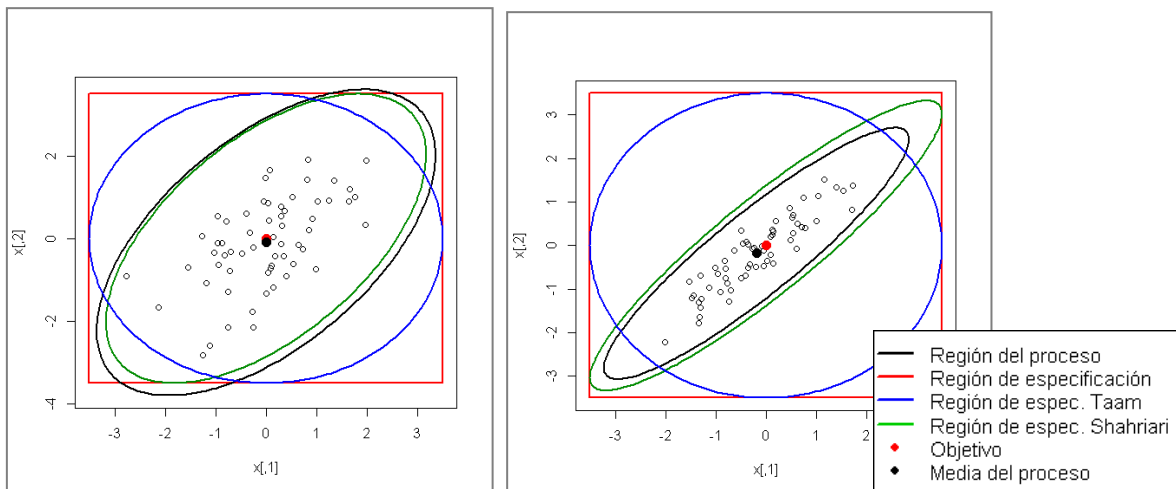


Figura 5: Valores medios de los indicadores de Taam, Shahriari y Wang según coeficiente de correlación (ρ) entre las variables.

Figura 6: Región de variación natural del proceso, región de especificación inicial, región de especificación modificada por Taam, región de especificación modificada por Shahriari, media del proceso y objetivo.



(a) Coeficiente de correlación

(b) Coeficiente de correlación



Algo similar ocurre con los enfoques basados en componentes principales. Mientras que el índice propuesto por Shahriari indica correctamente que el proceso bajo estudio es capaz, arrojando proporciones de productos dentro de especificación de más del 99%; el índice inicialmente propuesto por Wang genera subestimaciones de la capacidad real del proceso que aumentan a medida que la magnitud de la correlación entre las variables es mayor.

De manera análoga se ha analizado la gran variedad de escenarios considerados, cuyos resultados permitieron arribar a algunas conclusiones generales respecto del comportamiento de estos indicadores.

Se ha puesto de manifiesto que las propuestas iniciales en ambas líneas (Taam y Wang) no logran identificar de manera correcta el estado real del proceso, ya sea subestimando o sobreestimando la capacidad real del proceso teórico. Se ha observado además que ambos indicadores son inestables ante variaciones en el nivel de correlación entre las variables.

Por su parte, las propuestas más recientes logran mayor estabilidad ante variaciones en la magnitud de la asociación entre las variables. Entre ellas, la propuesta de los autores Shahriari et al. parece tener mejor comportamiento, en el sentido de identificar correctamente el estado real del proceso a través de la variedad de escenarios postulados. No sucede lo mismo con el indicador de Shinde, el cual si bien cuenta con la ventaja de ser fácilmente interpretable, no logra detectar correctamente ciertas situaciones de alejamiento del proceso respecto de las especificaciones.

Por otro lado, el análisis sobre vectores de variables de mayor dimensión, mostró que el aumento en el número de variables, potencia la inestabilidad y falta de capacidad de las medidas inicialmente propuestas para reflejar el estado real de los procesos.



Referencias bibliográficas.

SANTOS FERNANDEZ, E. y SCAGLIARINI, M. (2012). MPCl: An R package for computing multivariate process capability indices. *Journal of Statistical Software*, 47(7), 1-15.

SHAHRIARI, H. and ABDOLLAHZADEH, M. (2009). A new multivariate process capability vector. *Quality Engineering*, 21, 290-299.

SHINDE, R. L. and KHADSE, K. G. (2009). Multivariate Process Capability Using Principal Component Analysis. *Quality and Reliability Engineering International*, 25, 69-77.

TAAM, W., SUBBAIAH, P. and LIDDY, J. W. (1993). A note on multivariate process capability indices. *Journal of Applied Statistics*, 20, 339-351.

WANG, F. K. and CHEN, J. (1998). Capability index using principal component analysis. *Quality Engineering*, 11, 21-27.

WANG, F. K. and DU, T.C.T. (2000). Using principal component analysis in process performance for multivariate data. *Omega*, 28, 185-194.